Remarks

Applicant traverses the Examiner's rejections of the claims.

The claims of this application relate to the use of *indexes* and *statistics* in forming access plans for a database. An index is frequently used to access desired data in a database, and often statistics are used prior to executing a query, to estimate the likely size of the solution sets that will be generated from each selection criterion in a query. Typically, statistics are generated using an index in combination with the specific selection criterion of the query being processed.

For example, consider a car owners database, and a query seeking persons named "Smith" that live in the city of "New York" and own a "Packard" car. For this query, statistics for each of these criteria would be generated from an index, i.e., how many persons in the database are named "Smith", how many live in "New York", and how many own a "Packard" car. These statistics would likely show that the most efficient path to the desired result would be first seeking "Packard" owners, then selecting those in "New York" named "Smith". This would be far more efficient than first identifying "New York" city residents or persons named "Smith". The statistics, generated from the index, would thus identify the most efficient path to the answer.

After collecting statistics for a given query, those statistics may be cached for later re-use. For example, subsequent queries that seek persons in the city of "New York", could re-use the statistic previously generated for that same criterion. However, in order to re-use statistics, those statistics must be valid for the current criterion. The number of rows having a city name of "Brainerd" might be substantially less than the number having a city name of "New York". Thus, statistics generated for a first selection criterion like "New York" on a given attribute, typically cannot be reused for a second selection criterion such as "Brainerd", but rather must be re-validated by re-accessing the associated index.

Unfortunately, the time required to access indexes to generate statistics can be a substantial fraction of total query optimization time; thus, re-validation of statistics represents a substantial loss of efficiency in database processing.

This application relates to two concepts both dealing with statistics and indexes.

Specifically, claims 1-10 relate to the use of previously generated statistics generated for an attribute, on different queries. In processing a query including a selection criterion on one or more attributes of a relation, a prior statistic generated for a prior different selection criterion on the same one or more attributes of the relation, may be used in processing the query, even though the selection criterion differs. According to the claimed invention, the decision to re-use a criterion, is based upon a measure of the entropy of the one or more attributes of the relation. Specifically, if the entropy measure suggests that different attribute values generate similar numbers of hits, then the statistics need not be re-computed. In this way, the re-validation of statistics may be performed more efficiently.

In rejecting claims 1-10, the Examiner has relied upon the Chadha '146 and Jones patents. However, neither patent relates in any way to the invention being pursued by the claims. Chadha '146 discloses a data mining method in which attributes of a relation are combined, based upon various criteria (dimension reduction), so that data mining can be more efficiently performed. This method and analysis does relate tangentially to entropy and correlation of attributes, as "data mining" may involve entropy computations, as the Examiner has noted in col. 5 of Chadha '146. However, in no way does the method or topic of Chadha '146 relate to the main point of this application and its claims, which is the re-use of statistics created for one access plan, in another access plan. More specifically, Chadha '146 in no way relates to "revalidating a prior statistic generated for a prior different selection criterion", by the use of a "measure of entropy of [the] one or more attributes" of a relation. The Examiner's citation to text in Chadha '146 relating to dimension reduction, such as in column 7, in no way relates to generation of statistics in accesses to a database or the re-use of such statistics in subsequent accesses of a database.

The Examiner's citation of the Jones patent is equally inappropriate. Jones relates to accessing a database in response to a query. The text noted by the Examiner in column 7 of Jones states that statistical information is used to optimize a query plan. As such, Jones is no different than the background art noted above.

The Examiner has noted text in column 15 of Jones, which states that when data is to be sent to a receiver from the object server, the first step of performing this transmission is the "revalidation" of the request, followed by the formulation of a query plan. The Examiner appears to believe that the "revalidation" identified in column 15 is somehow descriptive of associating statistics for one selection criterion, to a second different criterion, as claimed. Applicant has been unable to find any support for such an interpretation of Jones. Rather, the "revalidation" described in column 15 appears to relate to the <u>request for or transport of</u> data that is to be sent to the receiver client, not to any statistics. Applicant notes that the term "validate" is used in Jones to refer to transport sessions, not to statistics, as seen for example at column 14, line 8. Furthermore, there is no apparent connection between the "revalidate" step 606 described in column 15, and the subsequent steps in which a "query plan is formulated. This query plan formulation process is analogous to that previously described." Rather, this text appears to indicate that query plan formulation is separate from the "revalidation" of the request, not part of it as the Examiner posits.

In short, Applicant finds nothing in Jones to suggest anything more than the use of statistics in the normal fashion, to create access plans. More particularly, there is nothing in Jones to suggest the claimed invention of "revalidating a prior statistic generated for a prior different selection criterion", by the use of a "measure of entropy of [the] one or more attributes" of a relation.

Turning to claims 11-24, these relate to a method for identifying groups of attributes for which a multi-dimensional index can beneficially be formed; this involves evaluating the correlation of attribute values within tuples of the relation, and determining that the correlation of attribute values within tuples of the relation exceeds a threshold. If so, then a multi-dimensional index can be usefully constructed over the correlated attributes.

The Examiner's rejection of these claims is based upon the Chadha '495 patent. Chadha '495 generally discloses the use of a particular index type, known as the encoded vector index, to index the content of a relation. The Examiner cites to text in column 4 of Chadha '495 describing an attribute as a field or column of a relational database, and to text in column 7 of Chadha '495 stating that statistics are generated from an index to optimize an access plan, and to text in column 9 stating how updating of an encoded vector index is handled when new tuples are inserted into a relation. All of this is relevant to an encoded vector index and its use, but Applicant can find nothing in it that relates to the claimed invention. Specifically, Applicant has found nothing in Chadha '495 that in any way relates to the claimed steps of "computing a correlation of attribute values" and "forming a multi-dimensional index for a group of attributes within tuples of the relation having a correlation of attribute values in excess of a threshold." There is simply nothing in Chadha '495 that describes the use of correlation values between attributes in the decision to form multi-dimensional indexes for attributes.

The Examiner's rejections of various dependent claims rely upon one or more of Chadha '495, Chadha '146 and Jones. For the reasons noted above, which need not be repeated, Applicant respectfully submits that none of these references are relevant to the claimed invention. Applicant therefore respectfully submits that all claims are allowable and requests early transmission of a Notice of Allowability.

If any petition for extension of time is necessary to accompany this communication, please consider this paper a petition for such an extension of time, and apply the appropriate extension of time fee to Deposit Account 23-3000. If any other charges or credits are necessary to complete this communication, please apply them to Deposit Account 23-3000.

Respectfully submitted,

Thomas W. Humphrey
Reg. No. 34,353

Wood, Herron & Evans, L.L.P.
2700 Carew Tower
441 Vine Street
Cincinnati, OH 45202-2917

Voice: (513) 241-2324
Facsimile: (513) 241-6234